



**QUEEN'S  
UNIVERSITY  
BELFAST**

## **Proteomic profile of cystic fibrosis sputum cells in adults chronically infected with *Pseudomonas aeruginosa***

Pattison, S. H., Gibson, D. S., Johnston, E., Peacock, S., Rivera, K., Tunney, M. M., Pappin, D. J., & Elborn, J. S. (2017). Proteomic profile of cystic fibrosis sputum cells in adults chronically infected with *Pseudomonas aeruginosa*. *European Respiratory Journal*, 50(1), 1-11. <https://doi.org/10.1183/13993003.01569-2016>

**Published in:**  
European Respiratory Journal

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

**Publisher rights**  
Copyright 2017 ©ERS.  
This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

**General rights**  
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**  
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

Proteomic profile of cystic fibrosis sputum cells in adults chronically infected with  
*Pseudomonas aeruginosa*

Sally H. Pattison,<sup>1</sup> David S. Gibson,<sup>1</sup> Elinor Johnston,<sup>1</sup> Samantha Peacock,<sup>2</sup> Keith Rivera,<sup>2</sup> Michael M. Tunney,<sup>3</sup> Darryl J. Pappin,<sup>2</sup> and J. Stuart Elborn<sup>1</sup>

<sup>1</sup> Centre for Infection and Immunity, Queen's University Belfast, Belfast, Northern Ireland, BT9 7BL, UK.

<sup>2</sup> Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.

<sup>3</sup> School of Pharmacy, Queen's University Belfast, Belfast, Northern Ireland, BT9 7BL, UK.

Corresponding Author: Dr Sally Pattison, Centre for Infection and Immunity, Medical Biology Centre, 97 Lisburn Road, Queen's University Belfast, N. Ireland, BT9 7BL, UK. Email: [s.pattison@qub.ac.uk](mailto:s.pattison@qub.ac.uk). Phone: (+44) 028 9097 2698

Take-Home Message: CF sputum proteomics detects high abundance of NET proteins and identifies proteins correlating negatively with %FEV<sub>1</sub>.

## **ABSTRACT**

Lung disease is the main cause of morbidity and mortality in cystic fibrosis (CF) and involves chronic infection and perturbed immune responses. Tissue damage is mediated mostly by extracellular proteases, but other cellular proteins may also contribute to damage through their effect on cell activities and/or release into sputum fluid by means of active secretion or cell death.

We employed Multidimensional Protein Identification Technology to identify sputum cellular proteins with consistently altered abundance in adults with CF, chronically infected with *Pseudomonas aeruginosa*, compared to healthy controls. Ingenuity Pathway Analysis, Gene Ontology, protein abundance and correlation with lung function were used to infer their potential clinical significance.

Differentially abundant proteins relate to Rho family small GTPase activity, immune cell movement/activation, generation of reactive oxygen species and dysregulation of cell death and proliferation. Compositional breakdown identified high abundance of proteins previously associated with neutrophil extracellular traps. Furthermore, negative correlations with lung function were detected for 17 proteins, many of which have previously been associated with lung injury.

These findings expand current understanding of the mechanisms driving CF lung disease and identify sputum cellular proteins with potential for use as indicators of disease status/prognosis, stratification determinants for treatment prescription or as therapeutic targets.

## INTRODUCTION

The lack or reduced presence at cell surfaces of fully functional cystic fibrosis conductance regulator (CFTR) in people with cystic fibrosis (CF) results in a build-up of dense, dehydrated mucus that promotes infection and thus, chronic inflammation [1]. Additionally, CF has been labeled a mucosal immunodeficiency syndrome because of the direct proinflammatory effects of CFTR dysfunction on the dysregulation of epithelial innate immunity and airway leukocytes [2]. As such, the pathophysiology of the CF lung is highly complex, involving numerous interactions between multiple human cell types and a diverse microbiota [3-5], as well as the inherent impairment of several cell types.

Proteomic studies of the CF airways have investigated protein abundance in epithelial cells cultured *in vitro* or obtained from animal models or clinical samples [6-13]. Furthermore, several studies have assessed the extracellular proteomes of CF sputum, bronchoalveolar lavage fluid (BALF) and bronchial epithelial cell secretions [14-19]. The main observations of these studies are summarised in the online supplement (Table E1). Although these studies have all contributed to our understanding of CF, investigations of single cell types, *in vitro* cultures and animal models have limited relevance to *in situ* CF lung pathogenesis. To date, the contribution of the global proteome of all cell types in the proportions occurring naturally in CF sputum has not been investigated. We hypothesized that proteins detected in sputum cells can have a detrimental effect on CF lung health, both through their effect on cell functioning and activities, and through their release into extracellular sputum fluid by means of active secretion or cell death.

We employed quantitative comparative proteomics to characterise the global activity and protein composition of cells in CF sputum from adults chronically infected with *P.*

*aeruginosa* to provide insight into potential mechanisms of lung disease and to identify candidate biomarkers associated with poor lung function in this population. By employing these methods we were able to identify, in clinical samples, novel cellular proteins and activities likely to be clinically relevant as mechanisms for tissue damage and disease progression.

## **METHODS**

### **Study participants**

Twelve adults with CF, aged  $26.8 \pm 5.98$  (mean  $\pm$  SD; range 18-37) were recruited at Belfast City Hospital on admission for treatment with intravenous antibiotics for an acute pulmonary exacerbation as defined by Fuchs et al. [21]. Seven were male and five female. All were chronically infected with *P. aeruginosa*. Their characteristics and treatment regimens are recorded in Table 1 and Table E2 (online supplement), respectively. Three induced sputum samples were collected from each CF participant at times of differing clinical status – untreated pulmonary exacerbation, within 24h of completing intravenous antibiotic treatment of exacerbation and 4-15 weeks after completion of treatment. Lung function was measured as FEV<sub>1</sub> % predicted prior to each sputum collection. Twelve non-smoking, healthy control participants, aged  $29.6 \pm 4.27$  (mean  $\pm$  SD; range 25-39) were recruited. Eight were female and four male. A single induced sputum sample was collected from each control individual. Ethical approval for this study was granted by the Office for Research Ethics Committees Northern Ireland (Ref: 09/NIR02/68) and informed consent was obtained from all subjects.

	Age	M/F	PEx (previous 12 months)	% FEV <sub>1</sub>	CFTR Genotype
CF1	23	F	2	35-36	F508del/ F508del
CF2	36	F	0	26-36	F508del/ F508del
CF3	31	M	1	81-102	F508del/G551D
CF4	18	M	1	57-69	F508del/ F508del
CF5	25	M	2	52-60	F508del/ F508del
CF6	37	F	0	51-62	F508del/ F508del
CF7	28	M	0	26-30	F508del/ F508del
CF11	29	F	1	49-75	F508del/ F508del
CF12	22	M	2	31-41	F508del/ F508del
CF13	19	M	1	55-77	F508del/ F508del
CF15	28	F	1	73-98	F508del/R117H
CF16	26	M	1	40-54	F508del/ F508del

**Table 1:** Characteristics of the CF cohort. % FEV<sub>1</sub> records the range measured over the 3 sputum collections.

## **Sample preparation**

For each sputum sample the cell population was harvested from mucus plugs, washed and proteins extracted. Briefly, approximately 1g of mucus plug was homogenised thoroughly with 9 ml sterile phosphate buffered saline and the cells pelleted by centrifugation. Following two further wash steps and cell straining to remove clumps, the cells were pelleted to remove all traces of supernatant. The pellet was then homogenised thoroughly in lysis buffer and subjected to repeated freeze-thaw and water bath sonication to ensure cell lysis. Insoluble debris was removed by centrifugation. All work was carried out at 4°C and in the presence of a protease inhibitor cocktail. Full method details are given in the online supplement.

## **Proteomic methods**

Protein digestion and iTRAQ labelling were performed as described previously [904 Scuoppo, C. 2012] and details are recorded in the online data supplement. Protein digestion with trypsin and labelling with 8plex iTRAQ Reagents (AB Sciex) were performed as described previously [22]. Six test samples and a reference sample were mixed for simultaneous analysis by 2D liquid chromatography MudPIT on line with an LTQ Orbitrap Velos (Thermo Electron) as described previously [22]. Further details of the methods employed for MudPIT are recorded in the online data supplement.

## **Protein identification**

Peaklists were generated by Mascot Distiller and protein identification was performed by the Mascot search engine against the UniProt database plus NCBI non-redundant database, taxonomy *P. aeruginosa*. Further details of the methods used in protein identification are recorded in the online data supplement. The mass spectrometry proteomics data have been deposited to the ProteomeXchange

Consortium [23] via the PRIDE partner repository with the dataset identifier PXD001985 and 10.6019/PXD001985. Western immunoblot and enzyme-linked immunosorbent assay (ELISA) were used to confirm the mass spectrometry identification and quantification data as described in the online data supplement.

### **Statistical analyses**

SPSS was used for all statistical analysis unless otherwise stated. iTRAQ sample labels were used to calculate each protein's abundance per sample relative to the reference included in each mass spectrometry run, thus enabling comparison of protein abundance between samples. Hierarchical clustering of sample protein profiles was performed by PermutMatrix [24] using Euclidean distance dissimilarity and McQuitty's linkage criteria. Prior to cluster analysis proteins within the dataset were tested for collinearity (Bivariate Spearman correlation coefficient  $>0.9$ ) and four proteins that correlated highly with others were subsequently excluded. Independent samples T-tests and Mann Whitney U tests were performed to identify differentially abundant proteins ( $p < 0.05$ ) between the CF and control cohorts. Ingenuity Pathway Analysis (IPA) was then employed to predict: (a) Canonical Pathway activity in both the CF and control consensus core proteomes (Fisher's Exact Test); (b) altered Bio Function activity in CF relative to the control group (Fisher's Exact Test); (c) Bio Function up- or down-regulation in CF (IPA Regulation z-score algorithm).

For proteins detected in  $>80\%$  of CF samples, their percentage abundance within each sample was estimated by the percentage exponentially modified Protein Abundance Index (% emPAI [25]) using MS Excel. For proteins of consistently high abundance ( $>1\%$  emPAI) in CF, % emPAIs were compared between the CF and control cohorts (Mann-Whitney U test).



Bivariate Spearman correlation analysis and univariate analysis were performed to identify proteins whose % emPAIs correlated with FEV<sub>1</sub> % predicted and to identify independent factors. A Benjamini-Hochberg multiple comparisons adjustment was performed for the correlation analysis (MS Excel) to control the false discovery rate at <0.05. For each of the proteins correlating negatively with FEV<sub>1</sub>, paired T-tests and Wilcoxon signed-ranks tests were performed to compare the relative abundance (sample:reference) of matched CF pre-treatment and <24h post-treatment samples to identify differential abundances associated with antibiotic treatment of exacerbation. Again, the false discovery rate was controlled at <0.05 by a Benjamini-Hochberg multiple comparisons adjustment.

## RESULTS

### Core signature proteome in CF cohort

In accordance with inclusion criteria, the CF cohort comprised adults with CF who were chronically infected with *P. aeruginosa*. The cohort's variation in % FEV<sub>1</sub> predicted (26-102%) and collection of 3 sputum samples from each individual (covering clinical stability, exacerbation and following antibiotic treatment of exacerbation) prevented proteome bias towards any particular clinical status within this population. In total, 2210 human proteins were detected including 119 which were common to all 48 samples, including those from healthy controls. A further 38 proteins were detected in all samples belonging to only one cohort suggesting low abundance or absence in the other cohort, although differential abundance between the two cohorts cannot be proven statistically. Twenty-one proteins were detected only in the CF cohort and 17 proteins were detected only in the control cohort. Table

E3 in the online data supplement records all proteins detected in all the samples belonging to at least one cohort.

The heatmap in Figure 1, which clusters individual sputum samples according to relative abundance levels (sample:reference) of consistently detected proteins, shows that CF and control samples largely cluster separately. The two exceptions, CF3.2 and CF4.2 (marked with an arrow in Figure 1), are CF samples collected just after completion of antibiotic treatment for a pulmonary exacerbation.

### **IPA Canonical Pathways in CF and control cohorts**

The 136 proteins detected in all control samples and the 140 proteins detected in all CF samples were treated as consensus core proteomes for the two cohorts. Qualitative individual analysis of these proteomes was performed by Ingenuity Pathway Analysis (IPA) to determine Canonical Pathways predicted by IPA as active in each cohort. Figure 2 shows the top 5 Pathways predicted in the control proteome and all Canonical Pathways predicted by to have a >5-fold increase in  $p$ -value in the CF proteome compared to the control proteome, indicating likely up-regulation in CF. Those predicted as most up-regulated in CF are involved in immune functions including neutrophil recruitment, rearrangement of the actin cytoskeleton, phagocytosis and T cell signaling.

### **IPA Bio Function analysis of the CF differential proteome**

Comparison of relative abundance levels (sample:reference) between the two cohorts (36 CF samples versus 12 control samples) was performed for the 119 proteins detected in all samples in order to identify cohort differences. 54 proteins were found to be less abundant and 36 more abundant in the CF versus control cohort (independent samples T-tests & Mann Whitney U tests,  $p < 0.05$ ; Table E3, online data supplement). Abundance ratios (CF:control) for all differentially abundant

proteins were included for IPA Bio Function analysis irrespective of the degree of difference. Table E4 (online supplement) records the Bio Functions predicted to be differentially expressed (Fisher's Exact Test,  $p < 0.05$ ) in CF with Regulation z-scores  $> 1$  or  $< -1$  (IPA Regulation z-score algorithm,  $p < 0.05$ ), and those predicted with a higher confidence level ( $p < 0.0005$ ), but with a less noteworthy or absent Regulation z-score (shaded rows).

Inflammatory Response, Cellular Movement, Cell Death, Cellular Growth and Proliferation, and Free Radical Scavenging were predicted by IPA as most altered in CF compared to the control cohort, with regard to the extent of regulation (Regulation z-score). This is largely supported by the data in Table E5 (online data supplement), which records Gene Ontology (GO) biological process classifications for the 38 proteins detected consistently in only one cohort, and the 6 proteins identified by Mann-Whitney U test as showing differential abundance between the two cohorts.

### **High abundance proteins in CF sputum cells**

As high abundance proteins have greater potential for directly impacting the lung environment, we identified proteins that were consistently detected at high abundance per sample in the CF cohort and determined if their abundance levels per sample differed significantly from those of the control cohort. Six proteins were found to each comprise  $> 1\%$  of total human cellular protein per sample (% emPAI) in  $> 80\%$  of CF samples. Comparison of the % emPAIs for these proteins between CF and control cohorts detected increased abundance of all six in CF (independent samples Mann-Whitney U test,  $p < 0.01$ ). Their median  $\pm$  IQR % emPAIs are displayed in Figure 4 along with the corresponding data for the control cohort. The Human Protein Atlas [26] identifies all as being either highly or moderately expressed in bone marrow

hematopoietic cells (50% of which are myelopoietic cells) and all except cathepsin G as moderately or lowly expressed in lung macrophages. No statistically significant difference in relative abundance level (sample:reference) for these 6 proteins was detected when comparing paired CF pulmonary exacerbation samples before and <24h after antibiotic treatment (Wilcoxon signed-ranks test,  $p<0.05$ ).

### **Sputum cellular proteins correlate with lung function**

In order to identify any associations between specific proteins and lung function, the relative abundance levels (sample:reference) of the 337 proteins detected in >80% CF samples were tested for correlation with lung function as indicated by FEV<sub>1</sub> % predicted. Negative correlations (Bivariate Spearman correlation analysis,  $p<0.01$ ; false discovery rate <0.05) were identified for the 17 proteins listed in Table 2, implicating them as potential biomarkers of CF lung disease. Positive correlations were also identified, but as the focus of this analysis was to identify potentially injurious proteins only the negative correlations are presented here. Alpha-1-antitrypsin and olfactomedin-4 were independent factors for FEV<sub>1</sub> % predicted (univariate analysis,  $p<0.01$ ).

UniProt Accession	Description	Gene name	Spearman correlation (FEV <sub>1</sub> )	Signif. (p-value)	CF pre- /post- abx
D3DSM0	Integrin beta-2 *	ITGB2	-.751	0.00000	1.114
P80723	Brain acid soluble protein 1 *	BASP1	-.732	0.00000	1.343
P02675	Fibrinogen beta chain **	FGB	-.732	0.00000	1.690
P01009	Alpha-1-antitrypsin	SERPINA1	-.674	0.00001	
P30048	Thioredoxin-dependent peroxide reductase, mitochondrial	PRDX3	-.646	0.00012	
J3QLC9	Haptoglobin *	HP	-.617	0.00028	1.288
P08246	Neutrophil elastase	ELANE	-.608	0.00036	
Q08722	Leukocyte surface antigen CD47 *	CD47	-.608	0.00037	1.115
Q6UX06	Olfactomedin-4 **	OLFM4	-.601	0.00011	1.620
P15144	Aminopeptidase N	ANPEP	-.540	0.00067	
F8VV56	CD63 antigen *	CD63	-.586	0.00068	1.338
C9JNR4	Transforming protein RhoA	RHOA	-.578	0.00082	1.354
P01023	Alpha-2-macroglobulin	A2M	-.543	0.00192	
O95498	Vanin 2 *	VNN2	-.540	0.00205	1.475
O94804	Serine/threonine-protein kinase 10	STK10	-.535	0.00233	
E7ER45	Maltase-glucoamylase, intestinal **	MGAM	-.523	0.00106	1.314
P08473	Neprilysin	MME	-.511	0.00146	

**Table 2:** Correlation between relative abundance level (sample:reference) and lung function (FEV<sub>1</sub> % predicted) for proteins present in >80% CF samples at false discovery rate <0.05. \* and \*\* indicate proteins with significantly different abundance

levels ( $p < 0.05$  and  $p < 0.01$ , respectively) in the CF cohort when comparing matched pre-antibiotic and <24h post-antibiotic treatment samples. Median ratios (pre-treatment : <24h post-treatment) are recorded in the final column.

## **Longitudinal analysis of antibiotic treatment of CF exacerbation**

Figure E1 (online data supplement) shows that total protein yield (mg per g sputum) consistently decreased for each CF participant following antibiotic treatment of exacerbation and that follow-up samples collected during clinical stability more closely resembled pre-treatment samples with respect to yield. Individual proteins identified as correlating negatively with FEV<sub>1</sub> % predicted (n=17) or consistently at high abundance in CF (n=6) were analysed to compare the relative abundance (sample:reference) of matched CF pre-treatment and <24h post-treatment samples. As shown in Table 2, paired samples T-tests (including Benjamini-Hochberg multiple comparisons adjustment of false discovery rate adjustment to <0.05) identified 9 proteins with decreased abundance ( $p<0.05$ ) following treatment of pulmonary exacerbation. Figure 1 shows that for 7 of the 12 individuals with CF, samples collected during clinical stability (labelled “.3”) clustered more closely with the matched pre-treatment sample (“.1”) than the post-treatment sample (“.2”).

## **ELISA validation of candidate biomarkers**

As high abundance proteins (Figure 4) have greater potential for directly impacting the lung environment, we chose to validate myeloperoxidase and lactotransferrin as candidate biomarkers. Quantification of our existing samples by commercial ELISA confirmed their consistently higher abundance in CF compared to healthy control sputum cells ( $p<0.05$ , Mann Whitney U test, Figure E2 in the online supplement). Additionally, this demonstrated that commercial ELISAs are a practical method for follow-up validation of candidate biomarkers in a larger independent cohort.

## DISCUSSION

In order to further elucidate the mechanisms of CF-associated tissue damage in adults with chronic respiratory *P. aeruginosa* infection, this study has for the first time specifically investigated the global proteome and functional activity of the combined population of cells present in CF sputum. As such, it uniquely studies the protein abundance of cells *in vivo*, thereby taking into account both their natural frequency in the sputum cell population and the effects of intercellular interactions. Thus, it delivers a high degree of clinical relevance to our findings. Moreover, to accurately characterise a consistent CF phenotype for adults chronically infected with *P. aeruginosa* and avoid bias due to patient individuality, we considered only those proteins detected in all samples within the CF and healthy control cohorts. Thus, our study is more stringent than previously published proteomic studies that either included proteins detected at high frequency in at least one sample [18] or used pooled samples [14-16].

In agreement with previous investigations of CF sputum or BAL [14, 15, 17, 18], we found the CF proteome to be largely distinct from that of the control cohort. IPA predicted that Canonical Pathways which affect or involve actin cytoskeleton rearrangement directed by the Rho family small GTPases, including RhoA, Rac1 and Cdc42 were over-represented in the CF samples. Thus, our *in vivo* data confirms previous *in vitro* observations linking CFTR deficiency with increased RhoA expression [27, 28]. This may be significant for CF lung health as impaired efferocytosis (phagocytosis of apoptotic cells) [29] and reduced expression of iNOS [28] resulting from CFTR deficiency have been reported as mediated by RhoA up-regulation.



Similarly to previous proteomics studies of cell-free CF sputum/BAL, GO and IPA Bio Function analyses identified up-regulation of the immune and inflammatory response [14-18]. In particular, we found increased immune cell movement and activation, and generation of reactive oxygen species. These observations fit with existing knowledge about the CF phenotype which is typified by chronic inflammation and neutrophilia. Interestingly, an IPA predicted network identified lipopolysaccharide as a likely trigger of inflammation. Cell death, growth and proliferation were also predicted as dysregulated in CF, in such a way that both death and proliferation of immune cells were decreased, while both were increased in non-immune cells. These observations further support the growing understanding of the CF phenotype. Tissue damage and cell death is rampant, while excessive proliferation of respiratory epithelial cells leading to squamous cell metaplasia and dysplasia is common in CF [30, 31]. Delayed neutrophil apoptosis and clearance is thought to contribute to chronic inflammation [2] and our study detected CF-associated differential abundance of matrix metalloproteinase 8, annexin I and nicotinamide phosphoribosyltransferase, which have been implicated in delayed apoptosis and clearing of inflammatory and immune cells *in vitro* [32] and *in vivo* [33, 34].

Compositional breakdown analysis of sample proteomes found S100-A9, the histone H2B family, lactotransferrin, histone H4, cathepsin G and myeloperoxidase to be consistently present in high abundance in CF and to each be significantly more abundant in the CF cohort than the control cohort. Importantly, all six proteins are known constituents of neutrophil extracellular trap (NET) structures, the release of which by neutrophils exerts antimicrobial activity through the degradation of virulence factors, bacterial killing and antifungal activity [35-37].

As our study focused on sputum cellular proteins we did not look specifically for NET structures, which are extracellular, and therefore do not know if they were present. However, NETs have been found previously in the CF airways fluid and the abundance of free DNA that is characteristic of NETs correlated with obstructive lung function [38]. As well as killing microbes, NETs have been found to directly induce human alveolar epithelial and endothelial cell death, with the predominant cytotoxic effect being attributed to histones and myeloperoxidase [39].

Individually, these six high abundance proteins may also have significance for CF lung pathology and several have previously been reported as up-regulated in CF sputum and/or BALF [14, 15, 18]. Importantly, differential abundance of myeloperoxidase in CF sputum fluid has been correlated with FEV<sub>1</sub> % predicted [17], and myeloperoxidase and calgranulin B have been identified in sputum fluid as CF-associated autoantigens [16]. Cathepsin G has been implicated in degradation of surfactant protein A, so reducing the innate pulmonary antimicrobial defence [40], while S100-A9 has been shown to induce MUC5AC production in human bronchial epithelial cells cultured *in vitro*, suggesting a potential contribution to pathological mucin hyperproduction *in vivo* [41]. In a study investigating trauma-induced lung injury, Abrams et al. reported that histones induced NET formation and myeloperoxidase release *in vitro*, while histone infusion in mouse models resulted in lung structural damage and neutrophil congestion [42]. Additionally, extracellular histones H4 and H3.3 have been implicated in asthma-related inflammation and remodeling, and in chronic obstructive pulmonary disease (COPD), respectively [43, 44]. Consequently, our consistent detection of these proteins in such high abundance in the sputum cellular fraction identifies a potential risk to lung health,

either through their active secretion and NET formation or following passive release into the sputum fluid via cell death.

A number of proteins detected consistently in CF sputum correlated negatively with lung function measured as FEV<sub>1</sub> % predicted. Their biological/clinical relevance and potential for causing injury are discussed more fully in the online supplement. Longitudinal analysis of patient-matched CF samples showed differential abundances for 10 out of these 17 proteins in the CF cohort following intravenous antibiotic treatment of a pulmonary exacerbation. Interestingly, all the changes in abundance resulted in levels closer to those in the control cohort. Previously, Sloane et al [17] observed that for 4/13 adults with CF, their cell-free sputum proteome profiles moved closer to those of healthy controls following antibiotic treatment of an exacerbation. However, the latter observations were made by comparison of 2-D gel images rather than implicating and identifying particular proteins. We also found that, although sputum total protein yield decreased in the CF cohort by completion of antibiotic treatment, it had reverted to pre-treatment levels by the clinically stable follow-up time point. This suggests that proteomic changes associated with antibiotic treatment may be temporary.

Overall, the proteins detected as negatively correlating with FEV<sub>1</sub> suggest that the degree of inflammation, neutrophil influx, protease activity and airway remodeling may all contribute to defining lung function. Alpha-1-antitrypsin and olfactomedin-4 were identified as independent factors negatively correlating with FEV<sub>1</sub> although this does not demonstrate causation of harm. Acute inflammation-associated protein alpha-1-antitrypsin has previously been identified as a biomarker of pulmonary disease in CF which can be predictive of response to treatment and prognosis [45]. As an inhibitor of neutrophil serine proteases, which contribute significantly to lung

destruction in CF, its negative correlation with lung function might seem surprising. However, previous studies observed that plasma deficiencies during inflammation were associated with less severe lung disease in CF [46]. No association between olfactomedin-4 and CF lung health has previously been reported. However, olfactomedin-4 has been shown to down-regulate proinflammatory responses to bacterial infection and reduce bacterial killing mediated by granule serine proteinases [47, 48], which could have the potential to negatively affect lung health.

There are a number of limitations to our study. As is typical for untargeted exploratory proteomic studies, we used highly sensitive shotgun technologies to obtain quantitative, high proteome coverage analysis of sputum cells as a novel target. The trade-off for such in-depth analysis is a relatively low sample size. Given the sample numbers practicable, we maximised power by restricting the study to adults with CF chronically infected with *P. aeruginosa*, who experienced pulmonary exacerbation requiring hospitalization for treatment with intravenous antibiotics. While CF sputum proteome responses to other microbial infections may vary, we considered this cohort most informative as 57% of the adult CF population is chronically infected with *P. aeruginosa* [49]. As this study investigated only adults the findings cannot be extended to children or individuals not chronically infected with *P. aeruginosa*. Although proteins reported as potentially damaging were detected as consistently abundant or correlating negatively with lung function in our study, this is not proof of causation of lung injury. Additionally, the limited sample size requires that any potential significance as biomarkers either of lung function or disease severity warrants validation in a larger independent cohort. As an initial stage in this validation process we used commercial ELISAs to confirm findings in our existing cohort for myeloperoxidase and lactotransferrin. Both proteins were considered as

candidate biomarkers due to their consistently high abundance in CF sputum, increased abundance in the CF compared to the control cohort and previous associations in the literature with potential for lung damage. Thus, we demonstrated that commercial ELISAs are a practical method for follow-up validation in an independent cohort. However, while cohort size and characteristics limit the scope of our study conclusions, the findings are of exploratory interest and may have future clinical value within a defined commonly encountered CF cohort. Moreover, the investigation of all types of cells in the proportions occurring naturally in sputum is novel and the approach adopted in our study could be informative for other respiratory conditions.

In summary, within the cohort studied, i.e. adults with chronic *P. aeruginosa* infection, we have detected a distinct signature for the CF sputum cellular proteome, and have used pathway and functional analyses to predict its potential clinical impact. In particular, our data highlight the prevalence and predominance of proteins that can be associated with NET structures and have been reported as linked to tissue damage. Additionally, we identified a number of proteins that correlate negatively with lung function, many of which have previously been associated with lung injury. Overall, these findings expand the current understanding of the mechanisms behind *P. aeruginosa* associated CF lung disease in adults and identify sputum cell proteins as candidate biomarkers for disease status, prognostic indicators or as stratification parameters for treatment prescription.

## **ACKNOWLEDGMENTS**

Sources of support: SHP was funded by the Cystic Fibrosis Trust Project Grant

PJ545. DP, SP and the proteomics work were part supported by the NIH/NCI Cancer Center Support Grant 5P30CA045508.

## REFERENCES

- 1 Boucher RC. Evidence for airway surface dehydration as the initiating event in CF airway disease. *J Intern Med* 2007;261:5-16.
- 2 Cohen TS, Prince A. Cystic fibrosis: a mucosal immunodeficiency syndrome. *Nat Med* 2012;18:509-519.
- 3 Cox MJ, Allgaier M, Taylor B, Baek MS, Huang YJ, Daly RA, Karaoz U, Andersen GL, Brown R, Fujimura KE, Wu B, Tran D, Koff J, Kleinhenz ME, Nielson D, Brodie EL, Lynch SV. Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS One* 2010;5:e11044.
- 4 Guss AM, Roeselers G, Newton IL, Young CR, Klepac-Ceraj V, Lory S, Cavanaugh CM. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *ISME J* 2011;5:20-29.
- 5 Delhaes L, Monchy S, Fréalle E, Hubans C, Salleron J, Leroy S, Prevotat A, Wallet F, Wallaert B, Dei-Cas E, Sime-Ngando T, Chabé M, Viscogliosi E. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community - implications for therapeutic management. *PLoS One*. 2012;7:e36313.
- 6 Bensalem N, Ventura AP, Vallée B, Lipecka J, Tondelier D, Davezac N, Dos Santos A, Perretti M, Fajac A, Sermet-Gaudelus I, Renouil M, Lesure JF, Halgand F, Laprévote O, Edelman A. Down-regulation of the anti-inflammatory protein annexin A1 in cystic fibrosis knock-out mice and patients. *Mol Cell Proteomics* 2005;4:1591-1601.

7 Roxo-Rosa M, da Costa G, Luider TM, Scholte BJ, Coelho AV, Amaral MD, Penque D. Proteomic analysis of nasal cells from cystic fibrosis patients and non-cystic fibrosis control individuals: search for novel biomarkers of cystic fibrosis lung disease. *Proteomics* 2006;6:2314-2325.

8 Carvalho-Oliveira I, Charro N, Aarbiou J, Buijs-Offerman RM, Wilke M, Schettgen T, Kraus T, Titulaer MK, Burgers P, Luider TM, Penque D, Scholte BJ. Proteomic analysis of naphthalene-induced airway epithelial injury and repair in a cystic fibrosis mouse model. *J Proteome Res* 2009;8:3606-3616.

9 Gomes-Alves P, Imrie M, Gray RD, Nogueira P, Ciordia S, Pacheco P, Azevedo P, Lopes C, de Almeida AB, Guardiano M, Porteous DJ, Albar JP, Boyd AC, Penque D. SELDI-TOF biomarker signatures for cystic fibrosis, asthma and chronic obstructive pulmonary disease. *Clin Biochem* 2010;43:168-177.

10 Ciavardelli D, D'Orazio M, Pieroni L, Consalvo A, Rossi C, Sacchetta P, Di Ilio C, Battistoni A, Urbani A. Proteomic and ionomic profiling reveals significant alterations of protein expression and calcium homeostasis in cystic fibrosis cells. *Mol Biosyst* 2013;9:1117-1126.

11 Ruaniyar N, Gupta V, Balch WE, Yates JR 3rd. Quantitative proteomic profiling reveals differentially regulated protein in cystic fibrosis cells. *J Proteome Res* 2014;13:4668-4675.

12 Lavallée-Adam M, Rauniyar N, McClatchy DB, Yates JR 3rd. PSEA-Quant: a protein set enrichment analysis on label-free and label-based protein quantification data. *J Proteome Res* 2014;13:5496-5509.



13 Jeanson L, Guerrera IC, Papon JF, Chhuon C, Zadigue P, Prulière-Escabasse V, Amselem S, Escudier E, Coste A, Edelman A. Proteomic analysis of nasal epithelial cells from cystic fibrosis patients. *PLoS One* 2014;9:e108671.

14 Gray RD, MacGregor G, Noble D, Imrie M, Dewar M, Boyd AC, Innes JA, Porteous DJ, Greening AP. Sputum proteomics in inflammatory and suppurative respiratory diseases. *Am J Respir Crit Care Med* 2008;178:444-452.

15 MacGregor G, Gray RD, Hilliard TN, Imrie M, Boyd AC, Alton EW, Bush A, Davies JC, Innes JA, Porteous DJ, Greening AP. Biomarkers for cystic fibrosis lung disease: Application of SELDI-TOF mass spectrometry to BAL fluid. *J Cyst Fibros* 2008;7:352-358.

16 Pedersen SK, Sloane AJ, Prasad SS, Sebastian LT, Lindner RA, Hsu M, Robinson M, Bye PT, Weinberger RP, Harry JL. An immunoproteomic approach for identification of clinical biomarkers for monitoring disease. *Mol Cell Proteomics* 2005;4:1052-1060.

17 Sloane AJ, Lindner RA, Prasad SS, Sebastian LT, Pedersen SK, Robinson M, Bye PT, Nielson DW, Harry JL. Proteomic analysis of sputum from adults and children with cystic fibrosis and from control subjects. *Am J Respir Crit Care Med* 2005;172:1416-1426.

18 Gharib S, Vaisar T, Aitken M, Park DR, Heinecke JW, Fu X. Mapping the lung proteome in cystic fibrosis. *J Proteome Res* 2009;8:3020-3029.

- 19 Peters-Hall JR, Brown KJ, Pillai DK, Tomney A, Garvin LM, Wu X, Rose MC. Quantitative proteomics reveals an altered cystic fibrosis in vitro bronchial epithelial secretome. *Am J Respir Cell Mol Biol* 2015;53:22-32.
- 20 Pattison SH, Johnston E, Gibson DS, Peacock S, Pappin DJ, Elborn JS. Proteomics analysis of the chronically infected CF airways [abstract]. *J Cyst Fibros* 2012;11:WS11.10.
- 21 Fuchs HJ, Borowitz DS, Christiansen DH, Morris EM, Nash ML, Ramsey BW, Rosenstein BJ, Smith AL, Wohl ME. Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. *N Engl J Med* 1994;331:637-642.
- 22 Scuoppo C, Miething C, Lindqvist L, Reyes J, Ruse C, Appelmann I, Yoon S, Krasnitz A, Teruya-Feldstein J, Pappin D, Pelletier J, Lowe SW. A tumour suppressor network relying on the polyamine-hypusine axis. *Nature* 2012;487:244-248.
- 23 Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, Dienes JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolomé S, Apweiler R, Omenn GS, Martens L, Jones AR, Hermjakob H. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 2014;32:223-226.
- 24 Caraux G, Pinloche S. PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* 2005;21:1280-1281.

- 25 Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. Exponentially modified Protein Abundance Index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 2005;4:1265-1272.
- 26 Uhlén M, Björling E, Agaton C, Szigartyo CA, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, Berglund L, Bergström K, Brumer H, Cerjan D, Ekström M, Elobeid A, Eriksson C, Fagerberg L, Falk R, Fall J, Forsberg M, Björklund MG, Gumbel K, Halimi A, Hallin I, Hamsten C, Hansson M, Hedhammar M, Hercules G, Kampf C, Larsson K, Lindskog M, Lodewyckx W, Lund J, Lundeberg J, Magnusson K, Malm E, Nilsson P, Odling J, Oksvold P, Olsson I, Oster E, Ottosson J, Paavilainen L, Persson A, Rimini R, Rockberg J, Runeson M, Sivertsson A, Sköllerö A, Steen J, Stenvall M, Sterky F, Strömberg S, Sundberg M, Tegel H, Tourle S, Wahlund E, Waldén A, Wan J, Wernérus H, Westberg J, Wester K, Wrethagen U, Xu LL, Hober S, Pontén F. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* 2005;4:1920-1932.
- 27 Perez A, Issler AC, Cotton CU, Kelley TJ, Verkman AS, Davis PB. CFTR inhibition mimics the cystic fibrosis inflammatory profile. *Am J Physiol Lung Cell Mol Physiol* 2007;292:L383-395.
- 28 Kreiselmeier NE, Kraynack NC, Corey DA, Kelley TJ. Statin-mediated correction of STAT1 signaling and inducible nitric oxide synthase expression in cystic fibrosis epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 2003;285:L1286-1295.

29 Vandivier RW, Richens TR, Horstmann SA, deCathelineau AM, Ghosh M, Reynolds SD, Xiao YQ, Riches DW, Plumb J, Vachon E, Downey GP, Henson PM. Dysfunctional cystic fibrosis transmembrane conductance regulator inhibits phagocytosis of apoptotic cells with proinflammatory consequences. *Am J Physiol Lung Cell Mol Physiol* 2009;297:677-686.

30 Voynow JA, Fischer BM, Roberts BC, Proia AD. Basal-like cells constitute the proliferating cell population in cystic fibrosis airways. *Am J Respir Crit Care Med* 2005;172:1013-1018.

31 Hajj R, Lesimple P, Nawrocki-Raby B, Birembaut P, Puchelle E, Coraux C. Human airway surface epithelial regeneration is delayed and abnormal in cystic fibrosis. *J Pathol* 2007;211:340-350.

32 Scannell M, Flanagan MB, deStefani A, Wynne KJ, Cagney G, Godson C, Maderna P. Annexin-1 and peptide derivatives are released by apoptotic cells and stimulate phagocytosis of apoptotic neutrophils by macrophages. *J Immunol* 2007;178:4595-4605.

33 Jia SH, Li Y, Parodo J, Kapus A, Fan L, Rotstein OD, Marshall JC. Pre-B cell colony-enhancing factor inhibits neutrophil apoptosis in experimental inflammation and clinical sepsis. *J Clin Invest* 2004;113:1318-1327.

34 Gueders MM, Balbin M, Rocks N, Foidart JM, Gosset P, Louis R, Shapiro S, Lopez-Otin C, Noël A, Cataldo DD. Matrix metalloproteinase-8 deficiency promotes granulocytic allergen-induced airway inflammation. *J Immunol* 2005;175:2589-2597.

35 Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, Weinrauch Y, Zychlinsky A. Neutrophil extracellular traps kill bacteria. *Science* 2004;303:1532-1535.

36 Urban CF, Ermert D, Schmid M, Abu-Abed U, Goosmann C, Nacken W, Brinkmann V, Jungblut PR, Zychlinsky A. Neutrophil extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against *Candida albicans*. *PLoS Pathog* 2009;5:e1000639.

37 Bianchi M, Niemiec MJ, Siler U, Urban CF, Reichenbach J. Restoration of anti-*Aspergillus* defense by neutrophil extracellular traps in human chronic granulomatous disease after gene therapy is calprotectin-dependent. *J Allergy Clin Immunol* 2011;127:1243-1252.

38 Marcos V, Zhou Z, Yildirim AO, Bohla A, Hector A, Vitkov L, Krautgartner WD, Stoiber W, Griesse M, Eickelberg O, Mall MA, Hartl D. Free DNA in cystic fibrosis airway fluids correlates with airflow obstruction. *Mediators Inflamm* 2015; 2015:408935.

39 Saffarzadeh M, Juenemann C, Queisser MA, Lochnit G, Barreto G, Galuska SP, Lohmeyer J, Preissner KT. Neutrophil extracellular traps directly induce epithelial and endothelial cell death: a predominant role of histones. *PLoS One* 2012;7:e32366.

40 Rubio F, Cooley J, Accurso FJ, Remold-O'Donnell E. Linkage of neutrophil serine proteases and decreased surfactant protein-A (SP-A) levels in inflammatory lung disease. *Thorax* 2004;59:318-323.

41 Kang JH, Hwang SM, Chung IY. S100A8, S100A9 and S100A12 activate airway epithelial cells to produce MUC5AC via extracellular signal-regulated kinase and nuclear factor- $\kappa$ B pathways. *Immunology* 2015;144:79-90.

42 Abrams ST, Zhang N, Manson J, Liu T, Dart C, Baluwa F, Wang SS, Brohi K, Kipar A, Yu W, Wang G, Toh CH. Circulating histones are mediators of trauma-associated lung injury. *Am J Respir Crit Care Med* 2013;187:160-169.

43 Calvo FQ, Fillet M, de Seny D, Meuwis MA, Maree R, Crahay C, Paulissen G, Rocks N, Gueders M, Wehenkel L, Merville MP, Louis R, Foidart JM, Noël A, Cataldo D. Biomarker discovery in asthma-related inflammation and remodeling. *Proteomics* 2009;9:2163-2170.

44 Barrero CA, Perez-Leal O, Aksoy M, Moncada C, Ji R, Lopez Y, Mallilankaraman K, Madesh M, Criner GJ, Kelsen SG, Merali S. Histone 3.3 participates in a self-sustaining cascade of apoptosis that contributes to the progression of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013;188:673-683.

45 Sagel SD, Thompson V, Chmiel JF, Montgomery GS, Nasr SZ, Perket E, Saavedra MT, Slovis B, Anthony MM, Emmett P, Heltshe SL. Effect of treatment of cystic fibrosis pulmonary exacerbations on systemic inflammation. *Ann Am Thorac Soc* 2015;12:708-717.

46 Mahadeva R, Westerbeek R, Perry D, Lovegrove JU, Whitehouse DB, Carroll NR, Ross-Russell RI, Webb AK, Bilton D, Lomas DA. Alpha1-antitrypsin deficiency alleles and the Taq-I G-->A allele in cystic fibrosis lung disease. *Eur Respir J* 1998;11:873-879.

47 Liu W, Yan M, Liu Y, Wang R, Li C, Deng C, Singh A, Coleman WG Jr, Rodgers GP. Olfactomedin 4 down-regulates innate immunity against *Helicobacter pylori* infection. *Proc Natl Acad Sci USA* 2010;107:11056-11061.

48 Liu W, Yan M, Liu Y, McLeish KR, Coleman WG Jr, Rodgers GP. Olfactomedin 4 inhibits cathepsin C-mediated protease activities, thereby modulating neutrophil killing of *Staphylococcus aureus* and *Escherichia coli* in mice. *J Immunol* 2012;189:2460-2467.

49 Cystic Fibrosis Foundation Patient Registry: 2015 Annual Data Report to the Center Directors. Bethesda, Maryland: Cystic Fibrosis Foundation; 2016.

**Figure 1:** Distribution and relative abundance of human proteins detected in sputum samples collected from CF and healthy control cohorts. A. shows that 119 proteins were found in all 48 samples, i.e. both CF and control, while a further 21 proteins were found exclusively in all CF samples and 17 proteins exclusively in all control samples. B. shows hierarchical clustering (using Euclidean distance dissimilarity and McQuitty's linkage criteria) of individual sputum samples according to the relative abundance levels for all 119 consistently detected proteins. CF and control samples largely cluster separately with the only exceptions being 2 CF samples (marked with black arrows) collected after antibiotic treatment for pulmonary exacerbation. The scale is a  $\log_2$  ratio of relative abundance.



**Figure 2:** Differential activity of Canonical Pathways in consensus CF and control proteomes. Based on the set of proteins detected in all samples within a cohort, Ingenuity Pathway Analysis software applied Fisher's Exact Test to predict the Canonical Pathways likely to be active within that cohort ( $p < 0.05$ ). The top 5 pathways predicted in the control cohort concern general metabolism, clatherin-mediated endocytosis signaling and mitochondrial dysfunction and these were detected with approximately equal probabilities in both cohorts. Pathways where there is a >5-fold increase in  $p$ -value in the CF proteome compared to the control proteome, indicating likely up-regulation in CF are involved in immune functions including neutrophil recruitment, rearrangement of the actin cytoskeleton, phagocytosis and T cell signaling.

**Figure 3:** Ingenuity Pathway Analysis predicted network for proteins detected in CF involved in Inflammatory Response and Cell Movement Bio Functions. The intensity of red or green shading indicates the degree of increased or decreased abundance in CF relative to healthy controls, respectively. A red outline with no internal shading indicates that the protein was detected exclusively in the CF cohort. The network identifies lipopolysaccharide (LPS) as a likely inflammatory trigger.

**Figure 4:** Percentage abundance estimates (% emPAI median  $\pm$  IQR) for proteins comprising >1% of total human proteins in >80% of CF samples. Human Protein Atlas (26) annotations for protein expression in terms of tissue location and localised abundance are recorded. H, M and L indicate high, medium and low protein expression. All proteins are either highly or moderately expressed in bone marrow hematopoietic cells, 50% of which are myelopoietic cells, and all except cathepsin G are moderately or lowly expressed in lung macrophages.